iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57425

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Congestion Detection with Sentiment Analysis

Pinak Wadikar[1], Dr. Rupesh C. Jaiswal[2]
[1, 2]*E&TC Department, SCTR's Pune Institute of Computer Technology, Pune, Maharashtra, India*

***Abstract:** This paper describes an approach to get insights into congestion on roads as well as get the general feedback of the state of transportation in a particular city and/or state. Google Maps gives you the location of the congestion but not the reason behind it. In this innovative approach, we try to extract the cause of the congestion as well as its severity with the help of sentiment analysis on Twitter feed. This data is especially helpful for authorities to arrange medical help, civic authorities for city planning as well as the common man who is new to a particular location.*

I. INTRODUCTION

Social media and twitter feed has become an important source of news and information in the current times. Tweets about a particular event are almost instantly seen on social media feeds. These tweets can be constructively analyzed to get a better understanding of the event under consideration. In this project, we have tried to collate tweets related to traffic congestions as well as civic works - which may or may not lead to traffic congestions – at a particular location. If a particular congestion is a result of a mishap, we expect the sentiment of the tweets to be highly unpleasant in which case, the tweets can be processed for more information on the situation. On the other hand if the congestion is due to some ongoing civic work then we expect the tweets to be either pleasant or unpleasant to a certain degree. These can be shared with civic authorities for as feedback or for further improvement. The classification of tweets is done based on the sentiment of the tweet.

II. TECHNOLOGY STACK

Tweepy: Python library for Twitter API
Natural Language Toolkit: For Sentiment Analysis
Named Entity Recognition (StanfordNERTagger): To extract subject, object, location tags

III. PIPELINE

Every tweet on the feed is initially filtered using Tweepy APIs. It is then preprocessed using stopwards and tokenized using NER. We then extract the sentiment. For demonstration purposes we have 3 types of sentiments – negative: -1, neutral: 0, positive: 1. Based on the sentiment, we take a decision which is to either share with a civic worker or with emergency services.

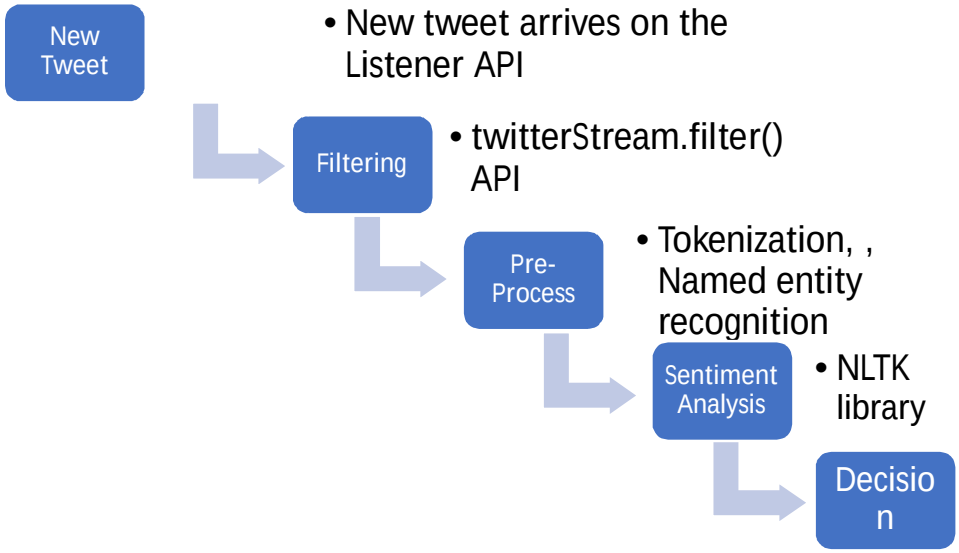


Fig.1.: Pipeline and processing of tweets

Fig.1 shows the processes through which a tweet is passed to extract sentiment as well as metadata

IV. GENERATING AND EXTRACTING TWITTER FEED

In order to demonstrate that relevant tweets are getting picked up by the API, we created dummy tweets on twitter with the necessary keywords:

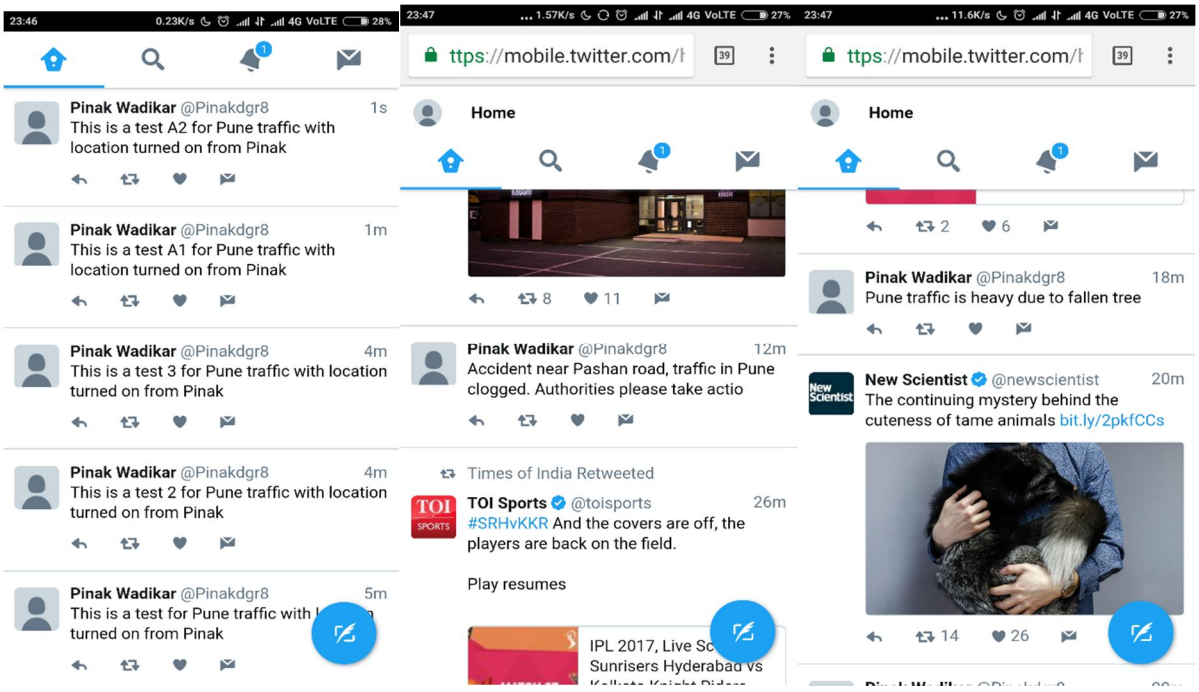


Fig.2.: Tweets from the author

Fig.2 shows the actual tweets on the twitter platform from the author of this paper. The tweets cover all possible scenarios. As you can see we have created tweets which show different sentiments.

Table 1: Tweet and Sentiments

Tweet	Sentiment
This is a test A1 for Pune traffic with location turned on from Pinak	0 (neutral)
Accident near Pashan road, Traffic in Pune clogged, Authorities please take action	-1 (negative)
No congestion at Nal stop. The new flyover is showing good results	1 (positive)

These tweets go on the twitter feed. We have selected few keywords with which we can pick up the tweets related to traffic and infrastructure.

Filter Keywords:

[‘Traffic’, ‘Congestion’, ‘road’, ‘flyover’, ‘Pune’, 'landslide','vvip','v.v.i.p','vip','v.i.p','vip','vips','crashed','speeding truck','truck','speeding','accident','toppled','overturned','overturn','tree','tree fall','speeding','truck"fall','fallen','landslides','blouder' ,'president','prime minister','accident','rammed','accidents','fallen tree','road', 'dug' , 'repair','men','work','work in progress','no entry','alert','blocked','trashed','landslides','landslide','prime-minister','vip','v.i.p','road repair','no entry','potholes','pothole','road cave','cave','road dug up','repair','fire','disaster on road','minister','rain','pipe','burst','construction','rain','downpour','flooding','raining','car','speeding','bike','rash','careless','driving','bus']

The filtered tweets can be seen here:

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

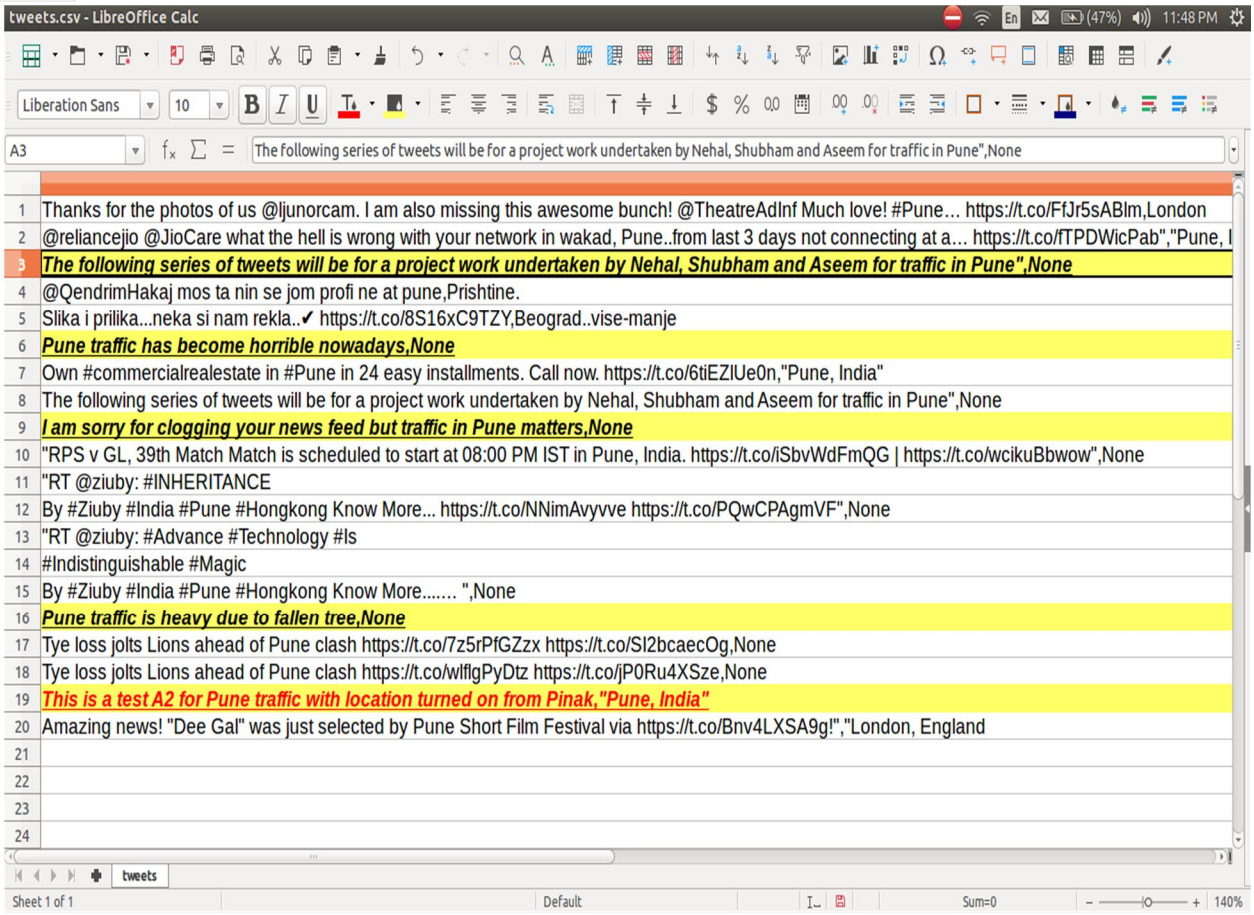


Fig.3.: Filtered Tweets

The shown csv file is constructed at run time. As you can see there are tweets about Pune from different users which are also collated in this CSV. These are not necessary for the project and will be filtered in the upcoming stages

As you can observe in the Filtered Tweets image, we do see many tweets with the keywords have been picked up by the Twitter listener. The highlighted are the ones created by the author (Pinak Wadikar). We do see that there are few other tweets which are also picked up by the API. These need to be removed which will be done in the Preprocessing stage.

V. PROCESSING THE TWEET (USING STOPWORDS ETC.)

The preprocessing stages is used for two purposes:

1) Extracting useful information from the filtered content

2) Tagging the information with Named Entity recognition.

For the advanced filtering of content, we make use of stop words. Stop words are commonly used words like prepositions ‘a’, ‘an’, ‘the’ that a search engine has been programmed to ignore. Along with the standard stop words from the NLTK corpus, we also added a few customized ones by analyzing the twitter feed that was given as an output of tweepy. Here is an example of stage 1 of preprocessing.

Filtered tweet: Pune Traffic is heavy due to fallen tree

Stage 1 processing output: [‘Pune’, ‘Traffic’, ‘heavy’, ‘fallen tree’]

The important words are then passed through stage 2 of preprocessing which is the named entity recognition. In our project the named entity recognition is particularly important because we need to extract information which is describing a location, structure or civic body. Thus, the tags ‘Location’, ‘Organization’, and ‘Person’ are important to us.

A sample output after the second stage of processing is shown below:

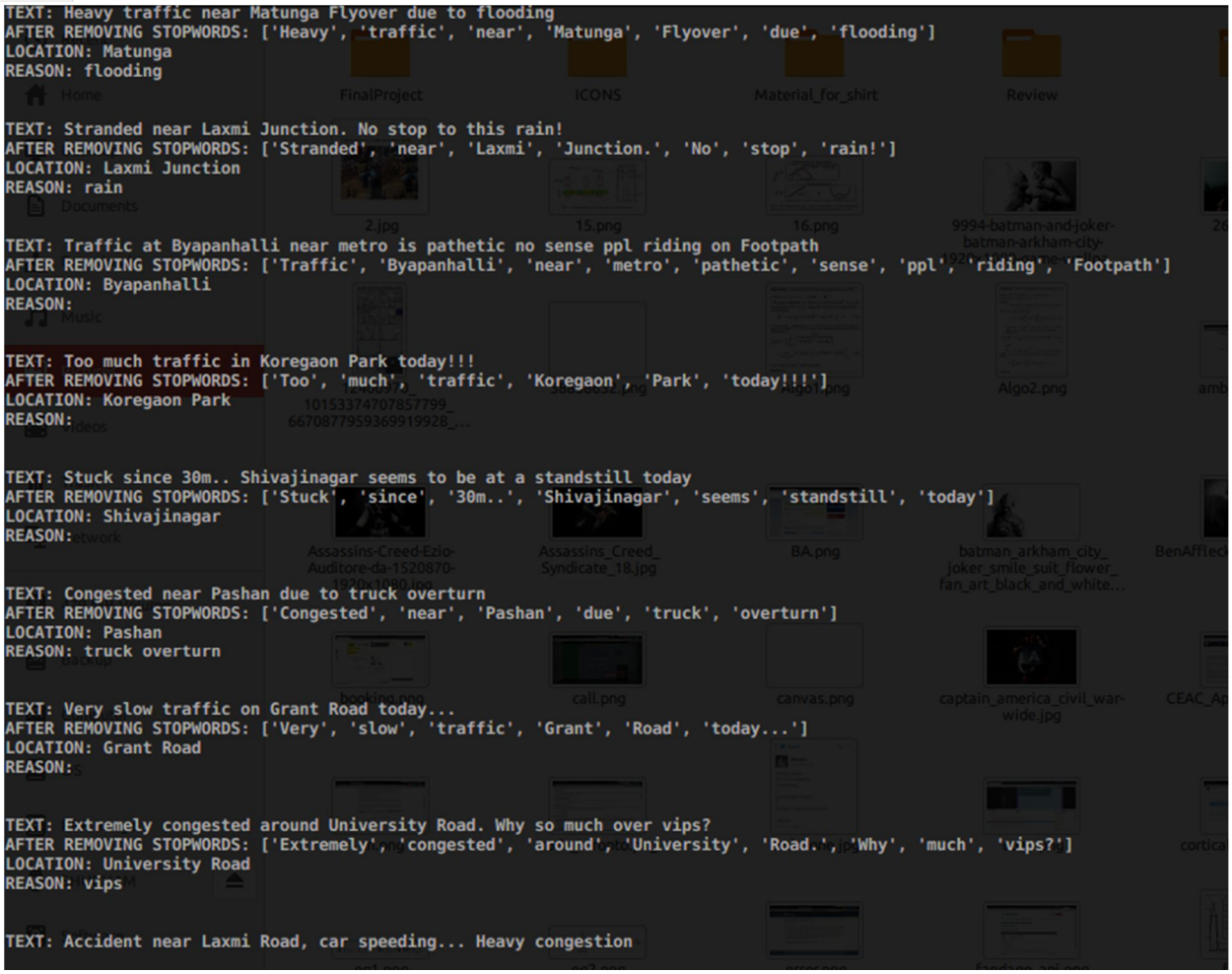


Fig.4.: Post processing output

The figure contains the tweets from which important information like the location and the reason behind the congestion have been extracted. This is not ready to be shared with authorities. The decision to share will be made after obtaining the sentiment of the tweet

VI. EXTRACTING THE SENTIMENT

A state-of-the-art Sentiment Classification system for social feeds data uses four methods for classification, namely – Bayesian Network, SVM, Decision Trees and Random Forest. This project required us to process data at high speeds while maintaining accuracy. There is a general tradeoff between speed and accuracy for the methods. The speed for decision trees is pretty fast however the accuracy is low. On the other hand, the Bayesian Network approach has high accuracy at the cost of performance. Thus, we decided to go with Random Forest classifier for sentiment analysis.

Random forest is a supervised ML algorithm based on ensemble learning. In the random forest algorithm, we construct a set of classifiers instead of a single classifier. It combines multiple decision trees to create a forest of trees. A voting system decides the final classification of the tweet.

Our classifier classifies data into 3 categories, namely:

-1: Negative

0: Neutral

1: Positive

Whether to share a particular tweet with authorities is decided based on the sentiment of the tweet. If the sentiment is negative, we get the output of the preprocess stage 2. The “Location” and “Reason” is extracted and shared with the authorities for quick action.

If the sentiment is positive, we collate such tweets and send them as feedback to authorities. A similar approach is taken for neutral sentiment.

Table 2: Tweet and Sentiment from Random Forest

Tweet	Sentiment from Random Forest
Heavy traffic near Matunga Flyover due to flooding	-1
Stranded near Laxmi Junction. No stop to this rain!	-1
Traffic at Byapanhalli near metro is pathetic no sense ppl riding on Footpath	0
Too much traffic in Koregaon Park today!!!	0
Stuck since 30m.. Shivajinagar seems to be at a standstill today	0
Congested near Pashan due to truck overturn	-1
Very slow traffic on Grant Road today...	0
Extremely congested around University Road. Why so much over vips?	-1
Accident near Laxmi Road, car speeding... Heavy congestion	-1
Heavy traffic near Pune University	0

VII. CONCLUSION

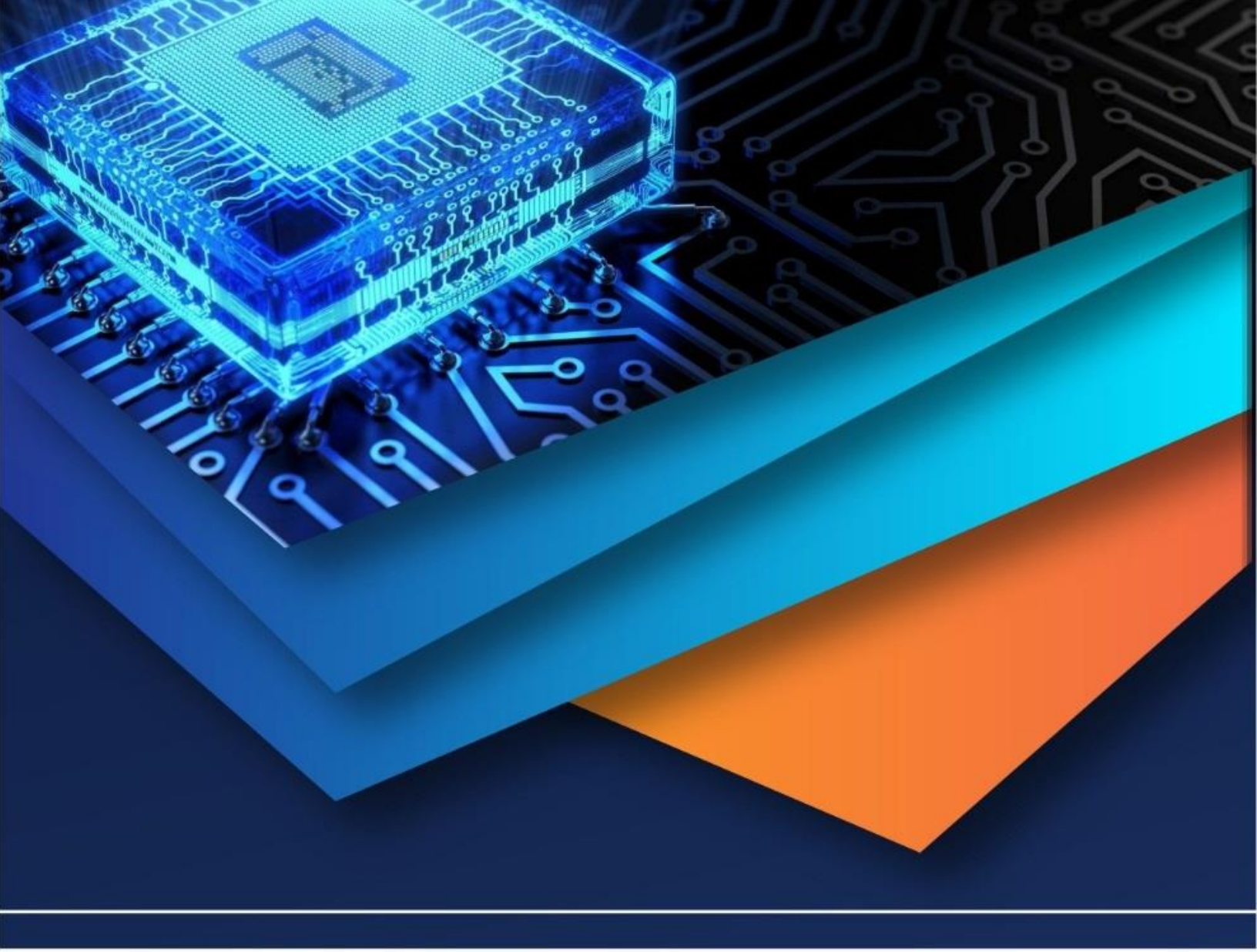
Congestion insights with social media feed was successfully implemented as part of this project. We demonstrated how important information from tweets could be extracted and shared with authorities for quick action or as feedback. We also demonstrated how sentiment of the tweet can be used to take emergency decisions. This implementation helps in prioritizing tasks for emergency services as well as town planning.

VIII. ACKNOWLEDGEMENT

I would like to express my gratitutde towards my mentor, Dr. R.C. Jaiswal for providing valueabl insights and guidance throughout the project. His perspective helped in expanding the applications of this project.

REFERENCES

[1] Bahrawi, “SENTIMENT ANALYSIS USING RANDOM FOREST ALGORITHM ONLINE SOCIAL MEDIA BASED”, December 2019
[2] Y. Wan and Q. Gao, “An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis,” 2015
[3] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, “Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Media,2011 , pp. 30-38
[4] R. Parikh and M. Movassate, “Sentiment Analysis of User- GeneratedTwitter Updates using Various Classi_cation Techniques",CS224N Final Report, 2009
[5] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, and Vasudeva Varma. 2017. Deep learning for hate speech detection in tweets. In Proceedings of the 26th International Conference on World Wide Web Companion, pages 759–760. International World Wide Web Conferences Steering Committee.
[6] Vedurumudi Priyanka, “Twitter Sentiment Analysis”, June 2021
[7] Jaiswal R.C. and Dhas Himanshu, “ Survey Paper on Stock Prediction Using Machine Learning Algorithms”, International Research Journal of Modernization in Engineering Technology and Science (IRJMETS), Open Access, Peer Reviewed and refereed Journal, Indexed in Google Scholar, Mendeley Advisor Community, ISSN: 2582-5208; Impact Factor:7.868, Volume 05 Issue IV, pp. 2744-2749, April 2023.
[8] Rahul Kulkarni, R. C. Jaiswal, “A Survey on AI Chatbots”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Open Access, Peer Reviewed and refereed Journal, Google Scholar, Mendeley : reference manager, Cite-Factor, Index Copernicus, ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:7.538, Volume 11, Issue IX, pp. 1738-1744, September 2023.
[9] “Tweepy”, an easy to use Python library for accessing Twitter API; https://www.tweepy.org/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY